Working Draft

United States Department of Agriculture

Service Center
Data Administration
Concept of Operations

August 26, 1998

Foreword

This document sets a basic framework for the duties, responsibilities, deliverables, and procedures which will make up the data administration process for the partner Service Center agencies. The partner Service Center Agencies include: the Natural Resources Conservation Service (NRCS), the Farm Service Agency (FSA), and the three mission areas within Rural Development.

The Service Center Data Team (Data Team) developed the material contained herein. This team is composed of representatives from NRCS, FSA, and Rural Development.

Sponsorship and direction for the Data Team comes from the Executive Director of the National Food and Agriculture Council (FAC) of the U.S. Department of Agriculture (USDA) and the partner Service Center agencies Chief Information Officers (CIOs).

Table of Contents

List of Exhibits

1 Introduction

1.1 The Business of Managing Data

In the 1970's and 1980's, the average cycle time for a new commercial business process was seven years. The average time to deploy a new information system was three years. In the 1990's, the cycle times have been shrinking, and the rate of this shrinkage is accelerating. In 1997, the average cycle time for business processes was only 12-18 months. A three-year effort to deploy information systems was no longer acceptable. Accelerated decision-making and streamlined development processes have started to reduce the information system deployment time. But further work is needed.

"Time" has become the *driving* variable in information systems design. "Time" has also become the *constraining* variable. A primary design goal for information systems should be to facilitate rapid changes in business and administrative processes. The design must look wider than immediate or apparent requirements, but not dwell on meeting all requirements for now and the future. The future is likely to be much different than expected.

The ability to change must be engineered "IN." Everything that prohibits change must be engineered "OUT." System development tools and application delivery methods are changing rapidly. The growth of networks is exponential. The response from industry in producing tools for rapid application development is dramatic. The one piece that is not progressing as rapidly is the collection and management of data.

Data is collected over a period of time, and often with a large expenditure of resources. Data is the foundation of any application system and its absence often impedes implementation of new processes. Existing data often outlives the application that created it, yet, more often than not, its value is overlooked. The capability for quicker development of applications is here, but the ability to collect and manage the data that fuels them is not.

Managing data is truly a business function. Data is collected and managed to support agency missions. Managers often seek the assistance of the Information Technology (IT) community in executing this function, but the responsibility for fulfilling the agency mission rests with the business area. The Data Team has been specifically assembled to provide the overall structure for good data management practices across all business areas, across agency boundaries, and across data users at all levels of the organization. The Data Team must work with the business areas and application developers to build a data management infrastructure which supports ongoing changes in business and systems, and which provides access to data needed for both national and local initiatives.

Anyone who handles data has a responsibility for maintaining its validity, and protecting it against improper usage.

Anyone who handles data has a responsibility for maintaining its validity and protecting it against improper usage. The lines of authority and responsibility for data may sometime seem blurred. However, if a person creates, stores, or uses data, they have a data management responsibility, regardless of the level of the organization.

The Data Team has coordinating responsibilities for data management within the partner agencies. The Team has set overall goals to make data available when and where it is needed, to manage and protect agency data assets, to promote a data architecture which can improve program delivery, and improve the return on investment in data collection. These goals are discussed in detail in a separate document, the "*Service Center Data Management Plan.*"

The ongoing Business Process Reengineering initiatives are emphasizing a data-centric approach to improving customer service. The Service Center objectives stress providing one-stop service by making data available through a single source, and reducing operating costs through data sharing.

The essential elements of managing data by business areas and information technology will become:

Essential Elements of Managing Data

- improving access to data at all levels of the organization and among agencies
- promoting data sharing among business areas and information systems
- reducing duplication in data collection; reengineering unnecessary data collection out of systems
- correlating individual data sets to expand the general knowledge base
- advertising the existence of data to a wider business audience
- facilitating common definitions of data for national applications among business areas, agencies, and levels of the organizations
- promoting methods to reduce application development times
- building a sound data infrastructure based on industry standards
- certifying and ensuring the integrity of data
- maintaining the security of data and the privacy of customers
- protecting the investments made in data assets
- establishing clear and concise standards and guidelines for database development and maintenance

1.2 Data Management and Data Administration

This document sets a basic framework for the duties, responsibilities, deliverables, and procedures needed to deliver an effective ongoing program of data management. It is useful at this point to explain the difference between Data Management and Data Administration, since the terms are often used interchangeably, but for our purposes are used to denote different functions.

Data Management is the managerial function of taking responsibility for data and the processes that support it. It focuses the strategic planning and operational data functions (technical planning, data administration, database administration, data warehouse administration) on meeting program delivery goals.

Data Administration encompasses the day-to-day technical functions that support ongoing business operations. It includes the collecting, defining, certifying, organizing, protecting, and delivery of both data and metadata (data about data).

The distinction is made because the responsibilities and duties of each function may be delegated and executed differently. The business area may hold some functions closely, while others are delegated. The Information Technology organization maintains a Data Management function that provides oversight, standards, procedures, and coordination that cut across business areas and agencies. IT Data Management usually would become involved in Data Administration functions only at the request of business areas, or to directly support the software application development centers, or where significant economies of scale can be demonstrated (such as in centralized database administration).

Exhibit 1 shows Data Administration as one of four ongoing data management reengineering initiatives. The Data Administration functions are outlined in this document. The Data Architecture, Data Warehousing, and Database Administration functions will be described in separate documents targeted at those specialized activities.

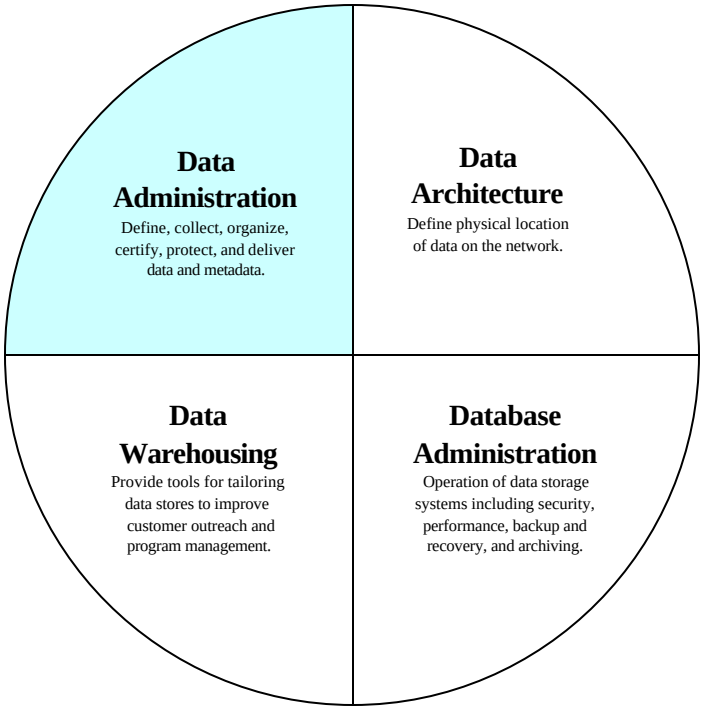

Exhibit 1 - Ongoing Data Management Initiatives

1.3 The Central IT Data Management Function

The Information Technology Data Management organization is responsible for coordinating overall data management activities across business areas. These responsibilities include such functions as acquiring and maintaining necessary software tools, managing a central repository for metadata, providing access to metadata for reporting and decision making, constructing an enterprise-wide view of data, promoting sharing and re-use of data assets across business areas, providing quality assurance over data management functions, managing data-related projects in support of one or more business areas, and other functions essential to a thorough data management process.

The Service Center Data Team (Data Team) is currently performing this centralized function. The Data Team is an interim interagency team established by the partner Service Center agencies. The team is responsible for implementing data management principles, policies, standards, and for establishing the overall data architecture. This Data Team will turn over its responsibilities to the data management organization within the new converged IT organization, when it becomes operational.

Many players are involved in managing data. These players are identified in Exhibit 2. They are grouped according to the enterprise-wide framework components that will be defined in Section 2 of this document. As shown, the activities of the Data Team will touch upon each of these components and players. Section 3 of this document identifies specific data management roles and responsibilities for both the data managers and the Data Team. Collaboration and cooperation are seen as the primary methods for implementing effective data management processes.

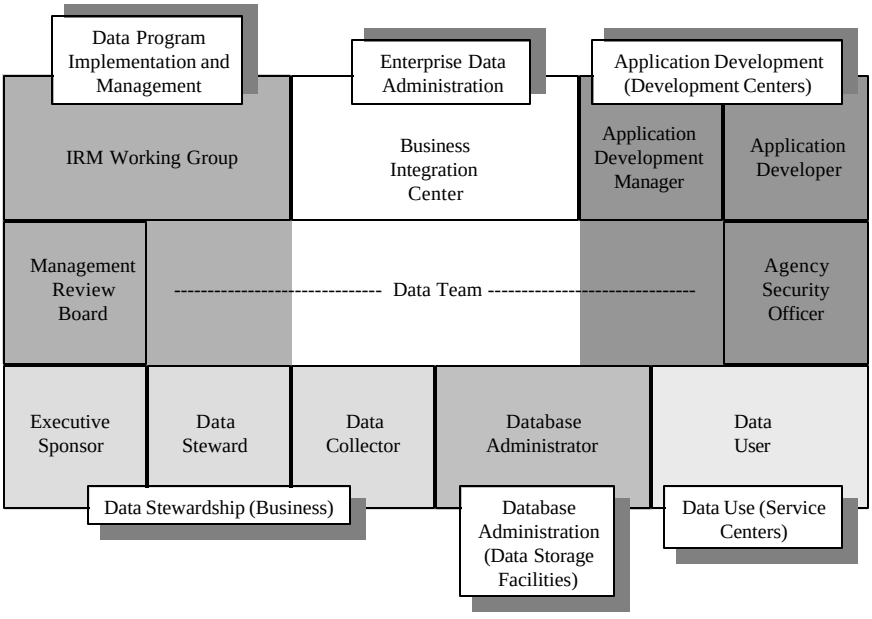

Exhibit 2 – Data Managers

1.4 National and Local Databases

The Data Management processes outlined in this document apply primarily to data and databases used nationally by the agencies. It is recognized that State and Local offices must maintain data to support strictly local situations and needs. These local databases are often created quickly to meet a specific situation, and usually kept for a short duration. While good management practices for the data in these local databases is expected these applications are not the focus of this document.

National and Local Databases are defined as:

Local Database

A database that is used by an agency organization (e.g., region, state, center, institute, etc.) but does not meet the definition of a National Database. Examples include local spreadsheets, report extracts, and PC-based databases.

National Database

A permanent database that (1) has international, national, USDA, or agency-wide application, (2) is included in a standard software suite, (3) contains data that is used/shared directly in making national program decisions, or (4) is used/shared in multiple offices, states, or other internal/external organizations.

1.5 Categories of Data

Because of the immense number and size of data stores held and used by the agencies, it is advantageous to group data according the potential breadth of usage or sharability among the partner agencies. Through Business Process Reengineering work already done, the following categories of data assets have been identified:

- **Common data** is data jointly owned, used, and managed by Service Center partners. The common datasets of interest for the BPR initiative are common customer data, office data, administrative data, common land unit data, and standard geospatial data.

- **Shared data** is data owned and managed by a specific Service Center partner and shared by other partners. The key shared datasets of interest for the BPR initiative are the natural resource datasets, specifically soils, plants, and climate.

- **Unique data** is data owned and managed by a specific Service Center partner but not shared.

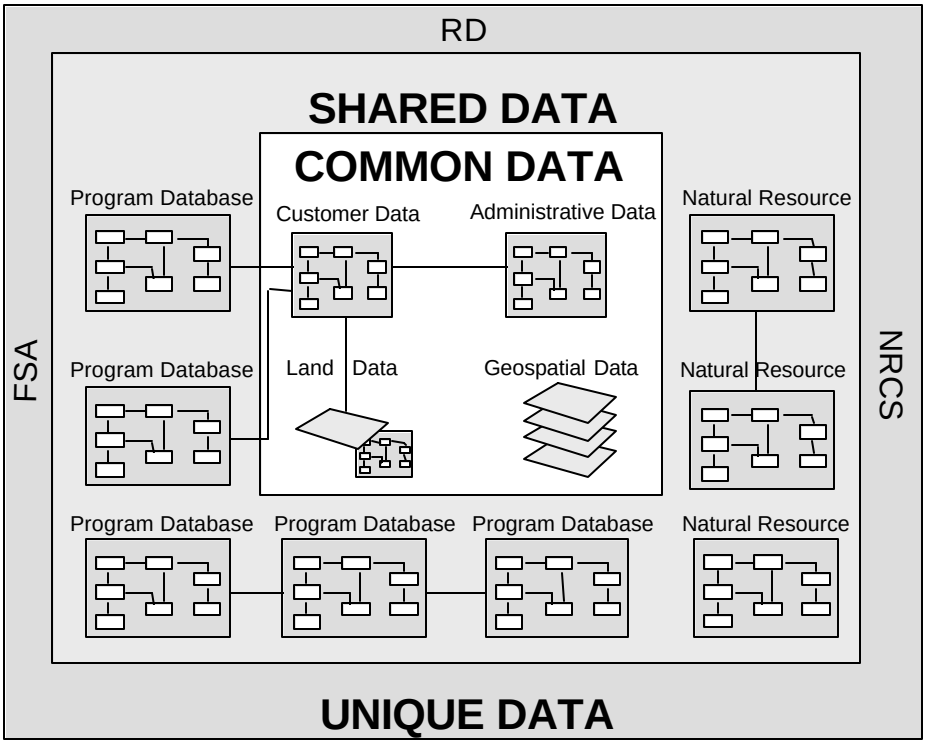

Exhibit 3 – Service Center Enterprise Data

The Data Team will initially focus on a core set of data which includes the "common data," plus key elements from "shared" datasets. It is expected that this core data will endure through multiple business cycles, and it is where the greatest demand for data sharing and access is being felt. Attention to the core set of data should offer the largest potential return on investment.

More rigor will be applied to the definition, collection, and maintenance of data and metadata for the core set. Individual agencies will be encouraged and assisted in applying similar rigor to their unique data. Emphasis will be placed on data for new and reengineered national systems, with legacy systems being addressed at the discretion of the business areas.

2 Framework for Enterprise-wide Data Management

An "Enterprise-wide Data Management Framework," shown in Exhibit 4, will provide structure to data management roles and functions. This framework organizes the enterprises' many data management activities into four major categories: those that provide direction to data management activities, and those that deal with implementation; and those that are administered centrally (although not centrally located), and those that are dispersed to multiple managers. Six functional areas of responsibility are described within these categories. The framework further describes the boundaries and dependencies between these functions.

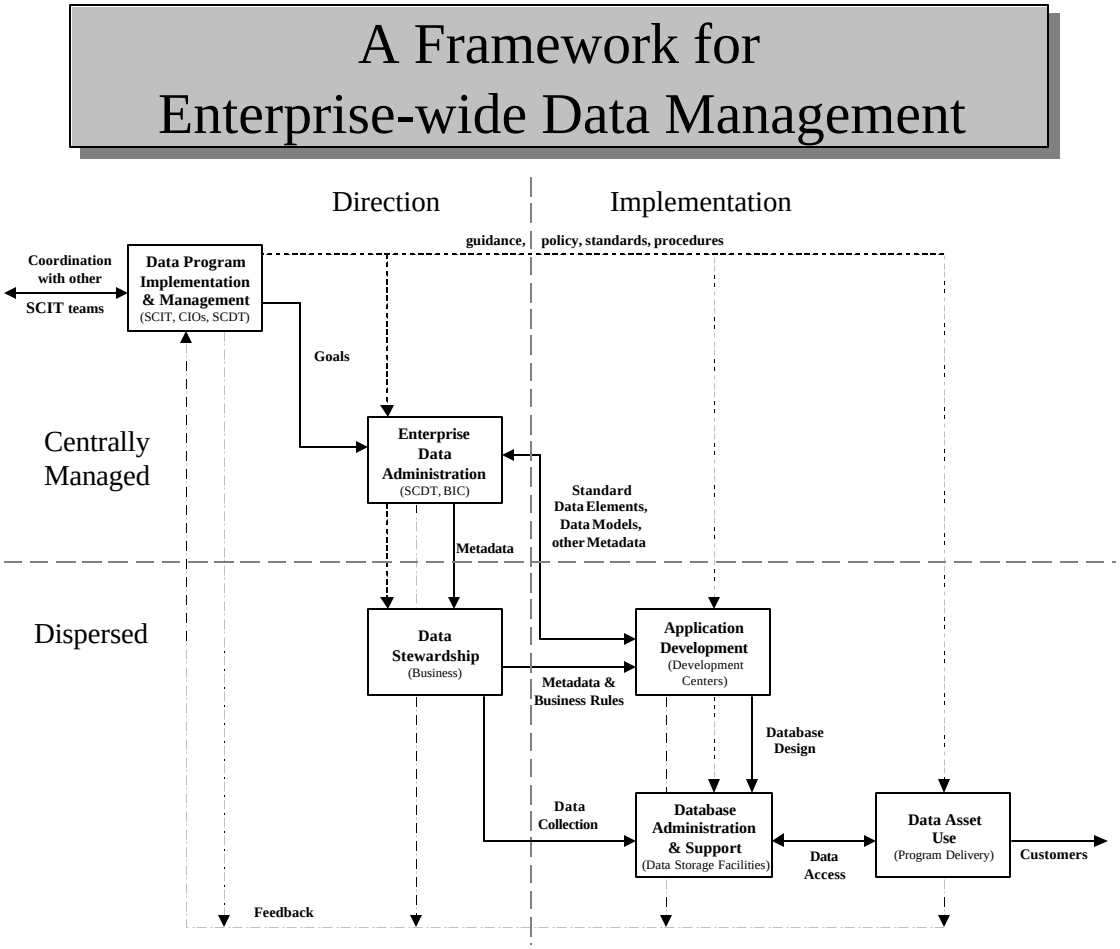

Exhibit 4 – Framework for Enterprise Data Management [1]

[1] This framework is adapted for the Normative Model of Enterprise Data Management that originated from work at MITRE in 1994 and 1995. This original work is presented in the Auerbach 1998 Handbook of Data Management. Burton Parker, one of the original contributors to the MITRE work, has evolved the original concepts from three to its current six components. This work is used by permission.

This framework addresses the intertwining roles of business areas with the enterprise-level and agency-level managers of data. The approaches outlined in this document will be constantly tested through the Business Process Reengineering initiatives, and will lay a foundation for the upcoming Information Technology Service Branch. A description of each of the six framework components follows. Additional detail on roles and responsibilities within each component can be found later in this document.

2.1 Data Program Implementation and Management

This component of the framework is responsible for the definition, coordination, and implementation of partner Service Center agencies' data administration activities. Primarily an agency and Service Center Implementation Team (SCIT) leadership role, this function establishes the data management vision and goals from an enterprise perspective. It establishes the organization, processes, and infrastructure to support enterprise data management activities.

A standing "Service Center Data Team (Data Team)", made up of representatives from FSA, RD and NRCS, will develop detailed data management policies, standards, and operating procedures. The members of this team need to be extremely knowledgeable of the service center business activities as well as the necessary supporting technology. All policies, standards, and operating procedures are reviewed by the partner agencies CIOs and the FAC Executive Officer (SCIT Team) who coordinate agency reviews and who function as a steering body for the Data Team. The Management Review Board (MRB) is responsible for final acceptance and implementation of data management policy.

2.2 Enterprise Data Administration

This component of enterprise data management framework is responsible for the on-going day-to-day coordination of data administration functions throughout the partner agencies (FSA, RD, and NRCS). It is responsible for identification, coordination, and organization of data that is potentially shareable across business areas, and with outside customers. The focus is on an enterprise view of data assets, with an objective to maximize data sharing and minimize redundancy in data collection.

Enterprise data administration entails defining the enterprise data architecture, managing an integrated enterprise data model, coordinating with business areas and software developers, maintaining data standards and conventions, coordinating the implementation of data policy, administering a central repository for metadata, identifying and managing standard data elements, providing data modeling assistance to projects, providing quality assurance oversight, and delivering training to those who create, manage and use the enterprise's data assets. These enterprise data administration activities are primarily carried out through the Service Center Data Team (Data Team) as they support the Software Development Centers and perform a coordination role in the Business Integration Center.

The Business Integration Center (BIC) is the focal point for all Business Process Reengineering (BPR) activities and particularly for the integration of business, data, and technical platforms. The BIC also provides a testing forum for cross-agency and agency-specific projects.

2.3 Data Stewardship

The data stewardship component of enterprise data management is a business-area function. Two key roles are "Executive Sponsor" and "Data Steward." An Executive Sponsor and a Data Steward are assigned by the business area for each database or set of data stored by the organization. Multiple people may be needed to support these operations, but the authority and responsibility must be defined.

The Executive Sponsor, usually the head of a division or other organizational component, has ultimate responsibility for the managing of data within the business area. The managing of resources (people and money) to collect and administer data and metadata is that person's responsibility.

The Executive Sponsor for each database/dataset may either retain the responsibility of Data Steward or appoint a Data Steward for that data. The Data Steward is responsible for the content of a database or dataset. The Data Steward establishes definitions and domains for data elements; sets the procedures for collecting and certifying data and metadata; and manages the overall storage, maintenance, and distribution of the data and metadata. This person often acts as a conduit between the end-user community and the IT community. The Data Steward may have frequent contacts with data administrators, application development teams, and database administrators supporting the business area.

A local Data Steward should be assigned at each point in the organization that data is stored. The same person can be assigned for multiple storage points, or it can be different people. Data that exists must have someone responsible for its use, security, and integrity. If data is widely distributed, stewardship must expand to cover it. Stewardship responsibility follows the data wherever it goes.

'Local Data Stewards' are delegated specific responsibilities by the Data Steward. These responsibilities may range from control and oversight of data collection and maintenance for an entire State to protecting a copy of a small subset of data used within a particular Service Center. Local Data Stewards should come from the business area staff at regional, state, area, or local offices. However, this responsibility may be delegated to someone outside the business area, including an Information Technology person or a data user. At local levels, for example, this responsibility may be assigned to a Service Center employee who is not directly part of the business area, but who uses data from multiple business areas in providing services to the public. All delegations of responsibility for data should be done formally, either by letter or through position descriptions and performance elements.

When copies or subsets of data are stored at state, regional, or county offices, at universities, or with local governments, the Data Steward and the Executive Sponsor need to establish 'Local Data Stewards' to maintain those datasets. Automation can be used to reduce the workload of Local Data Stewards in carrying out responsibilities related to use, security, and integrity needs of the data. Workload can be reduced but delegated responsibility is not diminished.

Data assets must always be within the span of control of the responsible business area, or their delegates, wherever this data is located. Copies of data may be released to outside parties, but only under conditions established by the business area to ensure the integrity of the data and the protection of privacy, data sensitivity, and security.

Collaboration and communication must take place among data stewards at all levels of the organization to ensure that data assets of the enterprise are protected, maintained, and used as intended. This is particularly true when data moves between machines and offices.

2.4 Application Development

Each application development project will be responsible for generating deliverables related to data. Examples of these deliverables include data models and related metadata. These products are essential to good system design, as well as being a reporting and coordinating mechanism. The core requirements will be initially established by the Business Integration Center.

As a minimum, application developers have a responsibility to ensure data used or produced by the system integrates with the enterprise data model, and that existing data is located and used where possible.

The majority of application development activity takes place at the partner agencies' software development centers located in Kansas City, Missouri; St. Louis, Missouri; Fort Collins, Colorado; and Washington, D.C. Application development projects will work with the data stewards to capture the business level metadata and develop data models. When development progresses into physical design, database administrators will be included to ensure physical database designs are efficient and tuned for the specific database and environment where the application will run. Throughout the development lifecycle, application development projects coordinate with the Data Team on enterprise level data administration activities.

2.5 Database Administration

This component of enterprise data management is responsible for supporting the physical or real instances of data stored within a database management system (DBMS). Activities include initialization, operation, tuning, maintenance, backup and recovery, archiving and configuration management of the database. Database administration support end-user access to the data assets so business activities can be accomplished. Database Administrators implement security rules within the database management system. They are integral to the identification of new data management tools and are responsible for the migration of data to new tools and environments. From the insight gained from managing day-to-day operations, database administration can provide valuable planning and feedback on issues relating to application performance, hardware capacity, etc. Database administration personnel may provide support remotely to multiple agency field locations.

2.6 Data Asset Use

This component of enterprise data management is the actual use of the data assets to carry out the business of the service centers, that is, the actual program delivery. Data assets exist to accomplish the agency mission and meet the needs of customers. This is therefore where the value of data assets is realized.

The benefits of the Secretary's "One-Stop Shopping" concept are realized here. Implementation of a data administration process that moves and stores data efficiently among offices, that allows ready access to central data stores, that integrates data across software systems, that correlates all

formats and types of data, and that provides searchable metadata for locating information pertinent to a customer or program are essential to reengineered customer service delivery.

Feedback from data asset users should determine whether the right information has gotten to the right people, at the right place, at the right time, and for the right cost.

3 Roles & Responsibilities for Managing Data

The effective use, protection, and maintenance of data assets requires that accountability and responsibility for data be assigned, and that adequate resources are identified to perform the functions. Primary roles are identified here, along with the responsibilities that go with these roles. *A single person can hold more than one of these roles. But, ultimately, all roles must be performed.* The "Framework for Enterprise-wide Data Management" described in section 2 is used for grouping these roles. The roles and responsibilities described here build on and further define the roles and responsibilities found in the data management policy document, "Interim Data Management Policy."

3.1 Data Program Implementation and Management

Management Review Board

The Board consists of the Chief Information Officers of the Partner Agencies, the Deputy Administrators for Management (DAMS), the Deputies for Programs, and the Executive Director of the National FAC. The Management Review Board will:

- Approve Service Center Data Management Policies.
- Provide coordination and means for implementing policies within agency organizations.

IRM Working Group

This Group consists of the Chief Information Officers of the Partner Agencies, and will:

- Recommend new Service Center Data Management Policies.
- Approve standards and procedures.
- Administer and ensure the enforcement of the Service Center Data Management policies and standards within their respective agencies.
- Approve waivers to Policy and Standards.
- Coordinate the review of policies, standards, and procedures with other organizational areas of the agencies that may be affected.
- Function with the SCIT Leader as a steering body for the Service Center Data Team (Data Team).
- Budget and provide staff and monetary resources to support operation of the Data Team.

Service Center Data Team Leader

The Team Leader is appointed by the Executive Director of the National Food and Agriculture Committee (FAC), and will:

- Establish and manage a multi-agency data administration function.
- Oversee the development of policies, standards and procedures needed to support data administration activities throughout the enterprise.

- Arbitrate in disputes or conflicts of data stewardship and usage among users of information and metadata.
- Coordinate with the FAC Director, CIO's, Agency leadership, Departmental staffs, and other governmental agencies.
- Manage the data coordination functions of the Business Integration Center.

3.2 Enterprise Data Administration

Service Center Data Team (Data Team)

The Team is managed by the Service Center Data Team Leader and executes the following functions:

Enterprise-wide Data Administration. This function includes:

- Establish a core Data Architecture, to include:
  - Maintaining the Enterprise Data Model for all new/reengineered applications;
  - Coordinating the collection of metadata for spatial and tabular data;
  - Maintaining the business rules supporting the Enterprise Data Model; and
  - Modeling the physical layout and location of National data used by the Agencies.
- Provide strategic planning for the acquisition and use of data assets to meet program goals.
- Provide management of the ongoing Data Management process, to include:
  - Review at a minimum annually and recommend modifications to the data management policies as appropriate;
  - Developing standards, procedures, and shared utilities and tools for data management;
  - Maintaining a shared Data Management Handbook that documents common data management standards and procedures; and
  - Coordinating implementation of a metadata repository, CASE, and modeling tools, and other supporting data management software.
- Provide a consolidated voice to the Department and to other government committees on data management issues.
- Establish a framework or structure for data administration processes.
- Institute key data management roles within the business areas, development centers, business process reengineering projects, and service center implementation organizations.
- Support development of data policies.
- Support the Business Integration Center and the BPR projects.
- Facilitate sharing and re-use of common data in Agency and Service Center applications.
- Implement a Change Control function for common and shared data.
- Provide quality assurance support to projects.
- Provide and implement Configuration Management standards and procedures for data stores and data management tools.

And such production data administration activities as:

- Perform Data Administration functions for applications, to include:
  - Resolution of conflicting data names, establishing common lookup tables, setting common domains for sharable data elements, and establishing unique keys and identifiers;
  - Coordinating data administration/management training; and
  - Coordinating the release of data to the public.
- Maintain a shared, central metadata repository for use by the Agencies to store and provide access to metadata, to include:
  - Establishing standard and sharable data elements to promote data reuse;
  - Making metadata accessible to system developers and other users; and
  - Maintaining a central source of data name abbreviations, and common acronyms.
- Coordinate and support Database functions, to include:
  - Coordination and review of logical and physical models;
  - Providing performance measures and metrics;
  - Implementing and coordinating security rules within the database;
  - Monitoring databases for performance; and
  - Coordinating and supporting reconstruction of databases to accommodate new information or to facilitate changes in the physical deployment of the database.
- Coordinate with security officers and database administrators to maintain data security.
- Coordinate with business areas and software developers on data administration issues.
- Maintain requirements for data tools, and coordinating the integration of tools.
- Provide direct data administration and modeling support to BPR projects and agency system development activities.

Data Warehouse Administration. In support of business functions, the Data Team will provide:

- Coordinate and support establishment of Data Warehouses to include:
  - Coordinating data that crosses business areas;
  - Coordinating the development and maintenance of the warehouse data model; and
  - Supporting the acquisition and management of data warehouse tools.

Business Integration Center (BIC) Data Administration. In support of the BIC, the Data Team will:

- Assist project teams in documenting project data according to standards.
- Review project deliverables.
- Provide quality assurance oversight.
- Coordinate posting of metadata to the central repository.
- Support Configuration Management activities.

3.3 Data Stewardship

Executive Sponsor

The Executive Sponsor is a business-area manager who is accountable for the collection, management, and use of data assets. The person has overall responsibility for the definition of the data, for the creation of software systems to collect and process the data, and all issues that deal with data content. In some cases this may be a shared responsibility between several business-area managers from different agencies.

The Executive Sponsor(s) for a National Database will:

- Determine if data, which the Agency plans to collect, has already been collected by the Agency, or whether cooperative efforts are possible to obtain the data from other existing sources (as required by Executive Order 12906). Establish cooperative agreements with non-agency sources of data.
- Coordinate funding for data collection, storage, and maintenance; and for software application development, support, and maintenance. Coordinate with partner agency management and other disciplines to set development and funding priorities.
- Promulgate and implement the policies and procedures necessary for ongoing management of the physical data content, the standards for the acquisition and certification of data, the policies for the collection and usage of metadata, and the procedures for the protection of the physical data assets.
- Designate a Data Steward, and other critical data management roles and responsibilities as appropriate. A data steward is assigned for each database or other set of data.
- Authorize the release of data and application software to internal and external customers.
- Have ultimate responsibility for the security of the data assets.
- Provide guidance and business-discipline support for the development and maintenance of application software necessary for managing the data.
- Certify that software applications meet discipline requirements.
- Change Management: manage "change" as it impacts the business discipline, the needs of customers, and the information delivery technology.

Data Steward

The Data Steward is a business-area expert who is assigned responsibility by the Executive Sponsor for the content of the database. Data steward responsibilities may be delegated to local data stewards who are responsible for portions or copies of a dataset. However, responsibility for the definition of the data cannot be delegated.

The Data Steward(s) for National Databases will:

- Act as the designated authority and point of contact for all business-area decisions concerning the database. Responsibilities include obtaining the needs/requirements from the users, and coordinating with the Data Team on metadata and other data management issues. Also, act as the point of contact for obtaining information on this data and for access to the data.
- Establish and maintain business rules and consistent definitions for data elements, identify data domains and relationships, establish data quality and certifications standards

associated with the contents of the database, and recommend availability, security and access authority for the data.
- Ensure that metadata is collected, approved, and certified for release according to adopted industry, Federal, and USDA metadata and data management standards. Also ensure metadata is made available according to the adopted standards.
- Ensure the validity, accuracy, and completeness of the physical data and supporting metadata; certify that data meets quality standards; and certify that data is ready for release for internal and/or public use.
- Provide training within the Data Steward's business area on data management roles and responsibilities.
- Provide guidance for the creation, storage, and dissemination of datasets.
- Develop technical procedures for acquiring and managing data.
- Implement quality assurance procedures for newly-collected and updated data.
- Identify training needs for data users.
- Coordinate with agency security officers to identify security requirements under the Freedom of Information Act, and for data that must be protected under the Privacy Act.
- Provide "help desk" support to governmental and outside users of data and supporting software.

Data Collector

The Data Collector is an agency staff member, or a contractor, who is assigned the task of gathering data for a particular geographic area, for the area served by an office, or for a particular business process (such as processing loan applications).

The data collector will:

- Gather data to meet the business area mission.
- Collect and document metadata required to describe data being collected.
- Ensure the accuracy of data and metadata to the extent possible.
- Ensure accuracy in the initial entry of raw data into automated systems and databases.

3.4 Application Development

Application Development Manager

This project manager is accountable for the development of information systems. The Application Development Manager will:

- Coordinate with the Executive Sponsor and Data Steward to develop information system requirements and design.
- Meet with the Data Team to develop a Project Data Management Plan.
- Provide the planning, coordination, staffing, and direction to create software, related databases and data stores, hardware and telecommunications platforms, and supporting services.
- Incorporate standards for ensuring data integrity into the design of the database and application systems.
- Ensure that the application is integrated into the enterprise model and data architecture.

- Follow the accepted systems development lifecycle, configuration management, and data management standards and procedures.

Application Developers

- Model the data used and stored by the application.
- Provide application-specific metadata to describe data generated and stored by the system.
- Ensure that existing data is used in the application wherever appropriate.
- Engineer the data created by the system to be sharable by other applications as appropriate.
- Coordinate with database administrators to ensure physical database designs are efficient and tuned for the specific database and environment where the application will run.
- Prepare the data management deliverables specified in the Project Data Management Plan
- Implement data security and privacy rules within the application.

Service Center Data Team

The Data Team, in support of application development, will:

- Coordinate data administration activities during the software development process and the system deployment phase. This will be accomplished primarily through Data Team members located at the development center.
- Create, or provide quality assurance for, data models, data dictionaries, and other metadata to support application development.
- Provide quality assurance and comments, as requested, on project proposals and designs.
- Ensure attention to the sharing and reuse of existing data.
- Integrate project information into the Enterprise Data Model.
- Establish "standard" data elements, abbreviations, acronyms, etc.
- Ensure project metadata is entered into the central repository.
- Provide analysis support for developing application data management procedures.
- Manage an interagency change control function for application data elements and data models.
- Coordinating and supporting reconstruction of databases to accommodate new information or to facilitate changes in the physical deployment of the database.

Agency Security Officers

A Security Officer is appointed by each agency to handle matters related to security, privacy, risk analysis, and disaster recovery issues for data, data systems, and telecommunications systems. The security officer will:

- Interpret Federal, USDA, and Agency policies on data security.
- Establish additional policies and procedures as needed to support applications and business usage of data.

- Coordinate with Executive Sponsors, Data Stewards, and Application Development Managers in defining and protecting sensitive data.
- Support database administrators in establishing system access, passwords, login ID's, and telecommunication access.
- Provide guidance in risk analysis, data retention, and disaster recovery procedures.

3.5 Database Administration and Support

Database Administrator

One or more Database Administrators may be appointed by the business area, or by the IT organization providing support to the business area. A database administrator will:

- Establish and maintain databases in support of applications.
- Determine the best physical organization and implementation for the databases.
- Tune the databases for performance.
- Support application managers and Data Team in planning databases within the Enterprise Model and Architecture.
- Coordinate with Agency Security Officers and data managers to implement data security and ensure the privacy of customer data.
- Implement security and access rules within the database.
- Provide for archiving, disaster recovery, and other long-term data protection procedures.
- Support the identification of new data management tools, and provide for the migration of data to new tools and environments
- Maintain currency of database management system software.
- Plan facility and equipment capacity to maintain access to the database.
- Provide planning and feedback on issues relating to application performance, hardware capacity, etc
- Implement necessary changes to the database structure across the organization.
- As needed, provide database administration support remotely to agency field locations.

3.6 Data Asset Use

Data User

This category consists of all persons that use the data assets, including service center staff, service center customers, partner organizations, state and local governments, outside users of agency information, members of the agency business areas, and IT management and staff. The data user will:

- Use data in the way it is intended to be used.
- Take responsibility for finding out the proper definition and usage of data.
- Provide information that allows data related to the user to be extracted and correlated.
- Take steps (security, login ID's, etc.) necessary to establish access to data stores.
- Provide thoughtful feedback to application developers and data stewards on the quality, utility, and timeliness of data.

4 Standards and Procedures

To support an effective data administration process a number of standards and procedures will be developed. The Data Team intends to use a document repository to manage their standards and procedures. Standards and procedures will be accessible via the World Wide Web.

- **Data Management Policy -** establishes the data management policy for the partner Service Center agencies. The standards and procedures developed by the Data Team are in support of the policy.

- **Metadata Content Standard –** defines the minimum metadata that projects must provide to populate the development center CASE tools and central repository.

- **Data Naming Standard** – a data element standard was completed earlier by the Data Team and is being actively maintained. Other naming standards will be developed.

- **Metadata Repository Interface Standard –** defines the physical interface between the central repository and development center CASE tools and other metadata sources.

- **Data Management Deliverable Standards –** a series of standards that define the content of data management deliverables for each phase of the system lifecycle.

- **Repository Administration Procedures –** defines the operating procedures for the central repository, including procedures for updating metadata and extracting standard data elements, procedures for security, maintenance and daily operations, and procedures for establishing standard data elements.

- **Enterprise Data Model Procedures –** defines the procedures for maintaining the Enterprise Data Model.

- **BIC Data Management Procedures –** defines the activities the Data Team will perform when software projects bring their deliverables to the BIC.

- **Common Data – Data Administration Procedures –** defines the data administration procedures for managing common data, common test data, and software components that can be shared.

- **CM/QA Procedures –** defines the CM and QA procedures needed to support data administration's CM and QA Plans, covering change control, review, audit and other processes.

5 Training and Project Assistance

The Data Team will provide training on data management policies, roles, processes, deliverables, and tools. The training will be available to both business and IT personnel, including managers, data stewards, project analysts, etc. A training plan will be developed to detail the training that will be provided.

The Data Team will also provide assistance to projects as they proceed through their development lifecycle. This assistance would include data modeling support, quality assurance support, assistance with impact analysis, and assistance with identifying data available for sharing or re-use.

6 Configuration Management

Data-related configuration management (CM) responsibilities include ensuring that versions of models and metadata are preserved for each production version of a system, that changes are adopted in an orderly and controlled manner, and that the status of models, metadata, and proposed changes is readily available. These CM responsibilities are administered at both the enterprise and the agency levels.

Configuration items of interest to data managers include:

- Tools – central repository and CASE tools, including the platform configurations required to run the tools
- Metadata contained in the central repository
- Datasets used for loading new databases, including common data and domains
- Enterprise data model
- Project level data models and metadata
- Standards, operating procedures, etc.

A detailed Configuration Management Plan will fully describe the CM deliverables and processes. The Data Management CM plan will be integrated with the SCIT CM Plan and Software Development CM Plans at the Development Centers.

7 Data Sharing & Integration

Two key responsibilities of the Data Team are to promote sharing and re-use of data across business areas, and to ensure that new data integrates into the enterprise model. The Data Team will assist projects in identifying existing data that can be incorporated into their project.

To further promote data sharing and integration, the Data Team will maintain a central repository of metadata and an enterprise data model. The Team will take an active role in the management of common data. Initial focus for the Data Team will be on the "30%" of data that is sharable.

7.1 Central Repository of Metadata

The Data Team will maintain a central repository to store, maintain, and distribute metadata. A detailed discussion of metadata can be found in Section 9 of this document. Emphasis will be initially placed on collecting metadata for the BPR projects, and in particular for common/shared data. The repository will also be available to support legacy systems and other application development initiatives as needed. The Data Team will provide interfaces for accessing the repository through CASE tools, through direct access tools, and via the web. They will provide the maintenance, operation, tuning, backup and recovery, and archiving of the repository necessary to make this a useful tool for application developers and other users.

7.2 Enterprise Data Model

An enterprise data model is a graphical and textual representation of all of the data used by an organization, how it is related to other pieces of data, the basic business function it supports, and who is responsible for its maintenance. The model serves as a device for planning and integrating the enterprise's data. The data model provides a view of the entire enterprise, its businesses and systems, from the perspective of the data vital to the enterprise.

The Data Team will maintain the enterprise data model. Each project's individual data model will be reviewed and integrated into the Enterprise Data model. Data will be grouped by subject areas. Integration points between subject areas will be emphasized. The enterprise metadata will be reviewed to ensure data ownership is maintained, key clashes are properly handled, and maximum sharing and reuse of data and related processes is attained. Linkage points with legacy systems that must be maintained will also be incorporated into the Enterprise Model.

Through the enterprise data model, the Data Team will monitor the use of record keys in various systems and business areas. The long-term objective is to establish globally-unique record keys for all common and shared datasets. Unique keys are essential if data is to move seamlessly between servers, laptops, service center computers, between offices, and over the web to customers. If a customer's records were to be accessed from anywhere in the enterprise, for example, a customer would need an identifier that is unique across the entire enterprise. In the same way, a method for uniquely identifying geographical features is needed if land unit data is to be readily accessible.

7.3 Data Administration for Common Data

Common data is data that is jointly owned, used, and managed by Service Center partners. This includes customer, administrative, land and geospatial data. The Data Team will provide direct data management support for this data, including providing CASE tools, data modeling support, and database administration support.

The Data Team will encourage the development of reusable software modules to encapsulate access to common data. These reusable software modules are referred to as components. All applications that need access to the data would do so via the provided components (see Exhibit 5). This approach offers a number of advantages. One of the key advantages is better control over integrity of the data. All changes to the data, i.e., creation, updates, and deletes, are handled by the components, allowing business rules to be implemented and enforced in one place. This relieves projects of the burden of writing their own access routines. Components also hide implementation detail, thereby reducing the impact of changes to the database. By ensuring that all access to data must pass through establish component(s), security and privacy is enhanced.

Components developed for accessing common data that can be incorporated into other systems, or called by other systems, will allow faster development of BPR projects. Components developed by an agency to provide access to "shared" data will help agencies protect their data stores while making them more accessible to users. The responsibility for maintaining the component and how it is "published" and made available to users must still be determined.

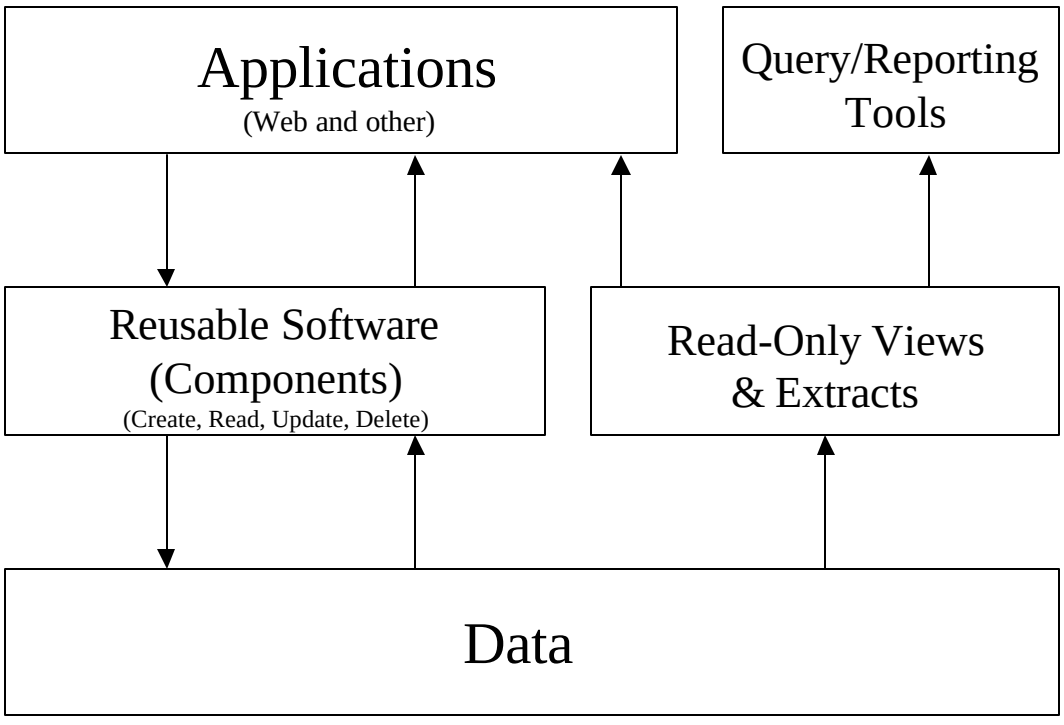

Exhibit 5 – Managing Data Access

8 Business Integration Center Process

Business Process Reengineering and other application development efforts will have contact with the Business Integration Center at specific points in their lifecycles. Data Management, along with LAN/WAN/Voice, CCE, and the Business Integration Center Coordinators, will have areas of interest to be addressed at each of these integration points.

Many of the Data Team activities discussed in this document exist to support the Business Integration Center. These activities are summarized here. The Data Team will provide content guidelines for the deliverables; standards for metadata, look-up tables, domains, and other common data structures; procedures for interacting with the Data Team at the Business Integration Center; configuration management standards; and change control and quality review procedures. The Data Team will also identify core requirements, such as minimum metadata requirements. The effort is to provide more structure to a set of products that should be a normal component of every project's documentation.

The data management activities and products, as they relate to the Business Integration Center, are detailed for the following project phases:

8.1 Initial Project Proposal

The initial proposal for a new application will contain a high-level business model. This model describes the business objectives to be attained, how this effort fits into the enterprise, what resources will be required, and what impacts the project will have. It is expected that the Initial Project Proposal will contain enough information to allow a decision on whether or not to continue the project.

Data Management is particularly interested in the data used by and created by a project, and the potential for data sharing and data reuse. All projects, whether a full software development effort or the implementation of a commercial software package, are expected to submit an Initial Project Proposal.

The Data Team will:

- assess the impacts of the project on the Enterprise Data Model and the Data Architecture
- integrate the high-level business model into the Enterprise Data Model

8.2 **Requirements** (a.k.a. the Reengineering Phase)

After a project is approved for further development, the Data Team will meet with the project manager and staff to develop a Project Data Management Plan. This plan is essentially an agreement with the project on the data management deliverables that make sense for that project.

The goal of this phase is to determine requirements of the reengineered business. Projects may provide such deliverables as conceptual and logical data models, data dictionaries, business rules, plans for data sharing and data re-use, a data migration strategy, and detailed definition of roles and responsibilities. Special attention should be made at this point to the use of existing data,

standard data elements, and the segmenting of functions into subsets that could become discrete, reusable components.

As part of the Project Data Management Plan, the Data Team can provide:

- training to developers on the data deliverables
- facilitation and quality assurance support
- detailed analysis on how the application will integrate into the enterprise
- refinement of estimates of project impacts on the technical platform, as system requirements are refined
- a metadata repository for project use, as well as management of that repository
- development support and quality assurance for the project's logical data model

Application requirements are to be reviewed by the Business Integration Center teams before moving on to the Design phase. This review process may include simulation testing to ensure anticipated impacts are acceptable.

8.3 Design

Design includes user interface design, design of the software architecture, physical database design, etc. Projects will provide data management information according to the Project Data Management Plan. Design information may include such items as a physical data model, data integrity rules, and complete metadata. For commercial software packages that are implemented as is, the physical design may be the only documentation needed for the package.

As defined in the Project Data Management Plan, the Data Team can provide:

- training on data deliverables
- consultation on data models
- impact analysis on the Enterprise Model
- quality assurance support

The technical design is to be reviewed by the Business Integration Center teams before moving on to the Development phase. Following successful review, the Data Team will:

- update the enterprise model
- check new information into the central repository
- place the data model under change control

8.4 Development

This phase includes development of the system, and project level testing done prior to handoff to the Business Integration Center. The Business Integration Center will perform final interoperability testing. Projects will adhere to the change control process for the data model and document all approved changes to metadata and business rules.

The Data Team can provide:

- quality assurance reviews
- impact analysis for changes
- test data for the common databases

The Data Team will:

- update the enterprise model for approved changes
- ensure approved changes are checked into the central repository

8.5 Integration Lab Testing

This activity consists of the final interoperability testing before a system is deployed as a pilot, national rollout, or upgrade. Interoperability testing will include an audit to ensure the metadata repository and the enterprise model reflect the as-built system submitted for testing. The database creation scripts are also placed under change control.

The Data Team can provide:

- test data for the common databases

The Data Team will:

- perform audit on data components of as-built system
- place database creation scripts under change control

8.6 Deployment & Maintenance

Once a system is deployed, the Data Team is available for assistance in problem resolution. The Team will participate in assessment of changes according to the change control process as defined in the CM Plan.

The Data Team can provide:

- assistance in problem resolution
- assessment and impact analysis on proposed changes
- development support and quality assurance for the as-built project data model

9 Managing Metadata

9.1 Definition of Metadata

"Metadata" is defined as data about data, or information about data. Three broad classes of metadata are 1) the information that describes the structure of data, and 2) the information describing physical instances of data, and 3) metadata that describes how data is aggregated.

The first class of metadata can be referred to as:

"registry" metadata - it gives data elements a name and definition, describes how the data will be stored, records relationships between data items, captures the business rules for data element and entities, describes the domain of valid data values, and identifies the caretakers of the data. It may include data models and database designs for specific application(s) where the data is used.

The second class of metadata describes the actual contents of data elements and data stores, and includes:

"clearinghouse" metadata - a physical instance of data that describes another set of data. Your driver's license, for example, is a physical instance of metadata that describes you, the person.

Clearinghouse metadata is used to describe and locate physical instances of data or software. Clearinghouse metadata can describe relational databases, data files, documents, photographs, video segments, stored sounds, software systems, applications, software components, libraries, etc. The level of granularity of the metadata may vary. In one case, the metadata may describe actual data objects. In other cases, the metadata may contain pointers to other sources or repositories of metadata. One set of metadata may list all available photographs and where to find them. The metadata describing each photograph in detail may be stored elsewhere with the actual photograph.

"GIS" metadata - this is tabular data that describes geospatial features. Points, lines, and polygons are created in a GIS tool and have little meaning if they are not described with words. These word descriptions can include such items as a name for the feature, category (i.e. farm, field, water, wetland), creation date, crop cover, ownership, etc. This is all tabular data that describes geospatial data, and is the lifeblood of GIS data.

There is GIS metadata, as well, that describes large sets of GIS data. This includes metadata that describes a complete map, or a geospatial data layer that covers a large physical area. This type of GIS metadata is used in particular to advertise or locate maps and GIS datasets.

A third type of metadata is often used to support data warehouses, and therefore is given that name:

"warehouse" metadata - this category can contain metadata describing information and data elements stored in a warehouse, similar to registry and clearinghouse metadata. Data used to populate a warehouse is usually imported from other production systems and databases. Warehouses, therefore, also require an extensive set of metadata describing how the data was brought into the warehouse. This information includes the original source of the data, timestamps, data conversion routines, data transformations, volatility of the data, refresh periods, data reliability indicators, relationships between data from multiple sources, etc.

The following example may help clarify the distinction between registry metadata and clearinghouse metadata.

Data vs. Metadata:

'043098,05101975' is data, but its meaning is unclear. Metadata would tell you that 043098 is an employee number, and 05/10/1975 is the employee starting date.

Registry Metadata vs. Clearinghouse Metadata:

A picture of the employee can be stored as a binary object in a database. Just pulling the picture up on a screen won't tell you who the person is. So we create a database element called employee number and populate it with the value '043098', and relate it to the picture. The value 043098 is data, but it is also metadata because it is information about another piece of data, i.e. the picture.

In this example, the data value 043098 is described by 'registry' metadata, which gives it a dataname of 'employee number'. In turn the data value 043098 is an instance of clearinghouse metadata, because it is a physical piece of data that describes another physical piece of data, the picture. We can search the database for all pictures related to employee 043098. But we can't search the picture data for all employees older than 25, unless we created another piece of clearinghouse metadata called 'birthdate'.

It is also useful to separate metadata according to its primary source:

Business-supplied metadata describes what is important to know about the data apart from a specific application or database implementation. It is information from a business or subject area perspective. Examples of this data asset include name, definition, format, ownership, business rules, etc. This metadata is usually supplied under the guidance of the Data Steward.

Application-specific metadata describes data as it has been implemented in a specific information system or database. It is information from an IT perspective – information needed to

develop and document the structures and inter-relationships of data. Examples of this metadata include data models, database designs, referential integrity rules, constraints, information about systems that create, modify, or use the data, etc. Application-specific metadata is usually supplied by the Development Center.

9.2 Metadata Uses

Metadata provides a source of knowledge about data that exists or will exist. Key users of metadata fall into the following categories: business areas, information technology, and outside customers. Business users and outside customers would likely have access to metadata through existing applications, reports, and warehouses.

Business Users include program managers, executive sponsors, data stewards, and system users. Program managers would be interested in definitions, glossary terms, etc. – information that helps them quickly understand information aspects of their business area or other business areas they need to interact with. Data stewards may need to do searches for existing data and associated software, and to gather information to support impact analysis. In addition, they are the primary providers of the business-supplied metadata.

Information Technology Users are those responsible for implementing software systems. This includes project managers, project analysts, data modelers, developers, database designers, etc. They will be interested in more physical aspects of the data and the software that accesses the data. When a system will be using common or shared data, they will need detailed information on how to access and use that data. The IT users will need information on physical formats, constraints, referential integrity rules, etc. Business-supplied metadata will need to be understood and implemented in specific databases and software. Metadata will provide a starting point for impact analysis by identifying what systems use and modify data. Metadata also enables sharing and reuse of data by providing identification, location, and stewardship information. IT is the primary provider and maintainer of application-specific metadata.

Outside Customers include other government agencies, universities, third-party software developers, and the general public. Their need for metadata falls into two categories 1) "clearinghouse" information to determine what data is available and how it can be acquired and used; and 2) "registry" information about the business meaning so fitness of use can be determined. "Customer" has broader scope than the customer supported directly by delivery of service center programs. There are a number of government directives that require that data, and information about the data, i.e. metadata, is made available to the general public. These directives include OMB Circular A-130, USDA Regulations on Public Access to USDA Electronic Information, OMB Circular A-16, Executive Order 12906, the Paperwork Reduction Acts of 1980 and 1995, etc.

In summary, many users will query the metadata to answer numerous questions. A small sampling of queries follow:

- What data do the agencies collect?
- Who owns the data (data steward/business contact)?
- Who should I contact for additional or specific information about the data?

- List all data elements in a specific piece of software.
- List all applications that use a particular data element/entity.
- Where is the data located?
- What physical format is the data in?
- What is the definition of a specific data element?
- Are there any constraints on the data, such as copyright or privacy act constraints?
- When was the data collected; when does it expire?
- How do I get the data?
- What are the valid values for a particular domain?

Some examples of metadata queries follow:

Scenario A: A local government agency wants to know what geospatial data themes are being collected by Service Center partner agencies.

- Define the subject area of interest – geospatial themes.
- Refine the search by limiting the search to themes of interest.
- Review the information on the theme, such as how it is attributed and what physical data stores or clearinghouses are available.
- Access clearinghouse for further information.

Scenario B: A data steward is searching for a source of customer data for a new BPR project.

- Define the subject area – customer.
- Review the attributes about customer to see if all needs are covered.
- Identify owner to contact to coordinate changes that will be needed for your project.

Scenario C: A project analyst needs to know the impact of proposed changes to physical data structures.

- Define the subject area of interest.
- Retrieve a list of systems that use, i.e., create, read, update or delete, data within the given subject area.
- Identify owners of those systems to contact for impact analysis.

9.3 Tools for Managing Metadata

As discussed, metadata is the information necessary for clearly describing, inventorying, analyzing, and classifying data. Data management policies, standards, and procedures provide the mechanism for ensuring that consistent metadata is collected, and that it is accessible by different users and organizations.

There are two main software tools that support the collection, management, and accessibility of metadata: Repositories and CASE tools. Exhibit 6 shows an overall approach for integrating a central repository with CASE tools to collect and access metadata. Exhibit 10 shows an overview for the collection of and access to geospatial metadata.

9.3.1 Repositories

A repository is a type of database tuned for capturing and maintaining specific types of information. Different types of repositories serve different purposes. Internal storage techniques and data search capabilities are tuned to support the types of queries needed in each case. Repository types include:

Data Registry – stores primarily tabular metadata that describes and defines individual data elements and data entities.

Clearinghouse – stores metadata about instances of actual physical data. The metadata may describe a data record, a map, a photograph, a spatial dataset, etc.

Warehouse – stores data from one or more sources in a way that facilitates management analysis of trends and patterns in the data. Also stored is the associated metadata describing how the data was brought into the warehouse, data conversions, data transformations, etc.

CASE Tool – stores the "CASE" information in a repository, sometimes called encyclopedias or dictionaries. Discussed in more detail in the next section.

Distributed Object Repositories – store interface and implementation metadata needed at runtime for locating and using distributed software components. These may also be referred to as registries.

Database Management System (DBMS) – stores metadata about the physical implementation of the data within the database. For a relational DBMS, this is typically referred to as a catalog.

The goal of a repository is to be the catalyst for unifying and integrating enterprise information resources. It facilitates collection of data and metadata in a consistent manner. When properly implemented, a repository will provide a single, authenticated source of data and its description serving one or more specific uses.

From the system developer's standpoint, a Data Registry can enhance communication and sharing of software system information across tools, lifecycle activities, users, and applications. Through this type of repository, the quality of system documentation can be maintained across development groups and project phases. Quality assurance procedures can be more consistently applied.

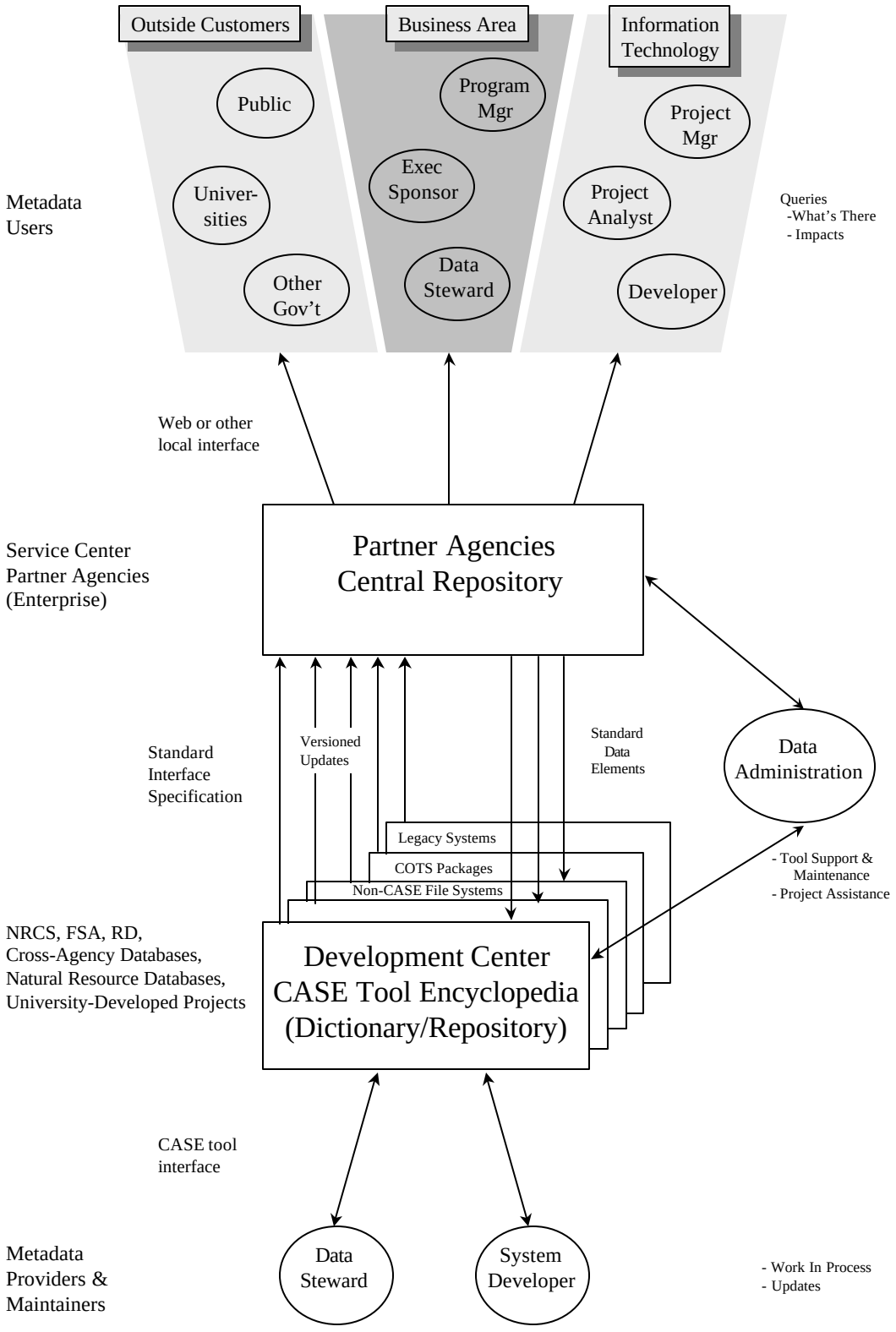

Exhibit 6 – Managing Registry Metadata

9.3.2 CASE Tools

CASE tools store their information in a tool-specific repository, typically referred to as an encyclopedia. These repositories may manage information at a single-user, multi-user, or multi-project level. However, they are typically proprietary solutions with closed architectures and cannot integrate information across the tools of the organization.

For CASE tools to provide real value within an organization, they must be part of an integrated environment. Specifically, they must be able to share and exchange information about data elements and data structures. Through integration with a central repository, this sharing and exchange of information can be achieved.

A diverse set of CASE tools is anticipated across development centers and even within a development center. A CASE tool's strengths and weaknesses are taken into consideration when determining which one best meets the needs of a development center or project. Since these needs will vary, a mix of CASE tools is likely. However, it is desirable to limit the proliferation of CASE tools. A list of sanctioned or preferred CASE tools will be maintained. Interfaces between sanctioned CASE tools and the central repository will be tested and supported.

Using the full facilities of a given CASE tool to provide an integrated environment for a particular application development project is an efficient scenario. CASE tools can then be chosen and tailored to the project and the technologies involved. Data models and metadata are created and updated within the CASE tool's workspace and stored in a CASE-tool specific encyclopedia. Check-out and check-in features allow work to be done on subsets of the project by multiple users.

When CASE Tools are used, the central repository will function as a passive repository, that is, updates are primarily done at specific checkpoints in the project lifecycle. These checkpoints follow major milestones in the project, with some intermediate deliverables to enhance data sharing among projects being developed in parallel.

The interface between these CASE tools and the central repository will follow defined interface standards. These interfaces can be specifically purchased for a given CASE Tool/Repository combination (Exhibit 7), or implemented by generic vendor-neutral tools (Exhibit 7). Examples of CASE tools with established interfaces include Sterling's COOL:Gen™ and Powersoft's PowerDesigner DataArchtitect™ when used with the MetaWorks™ component.

Another configuration (see Exhibit 8) is for projects that use the central repository as their primary work-in-process repository instead of a CASE tool. Special rigor must be used to ensure that data going directly into the repository is valid since no front-end system is managing the process.

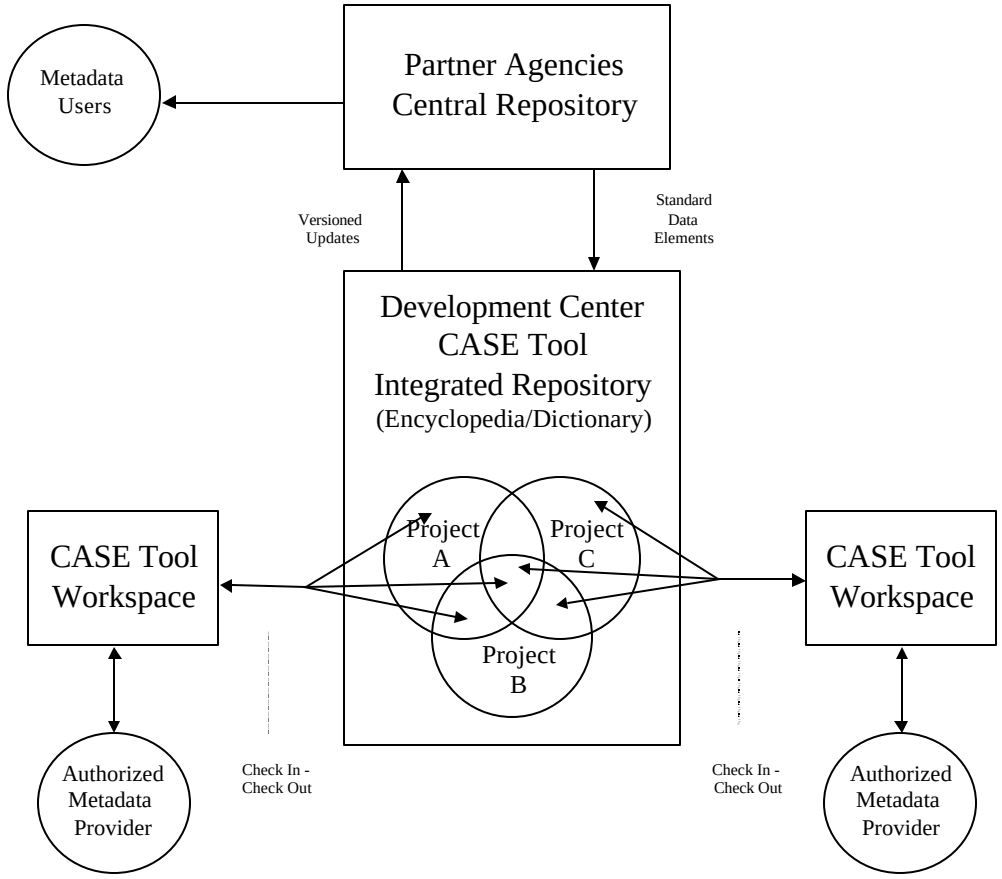

Exhibit 7 – CASE Tool with Integrated Repository

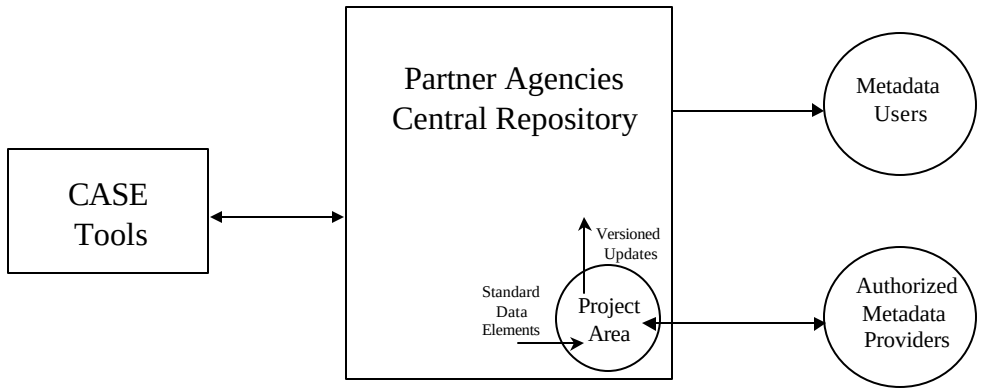

Exhibit 8 – Project Uses Central Repository As CASE Tool

Metadata from other sources can be entered into the central repository via a vendor-neutral interface (see Exhibit 9). These include:

- **COTS Products –** When COTS products are used, associated metadata will be obtained from the vendor. The level or amount of metadata may vary from product to product.
- **Legacy Systems –** Metadata may be collected from legacy systems using reverse engineering tools that deposit the information directly into the repository.
- **Non-CASE File Systems –** Some projects may use office automation tools such as word processors, spreadsheets, local databases, etc. to capture their metadata.

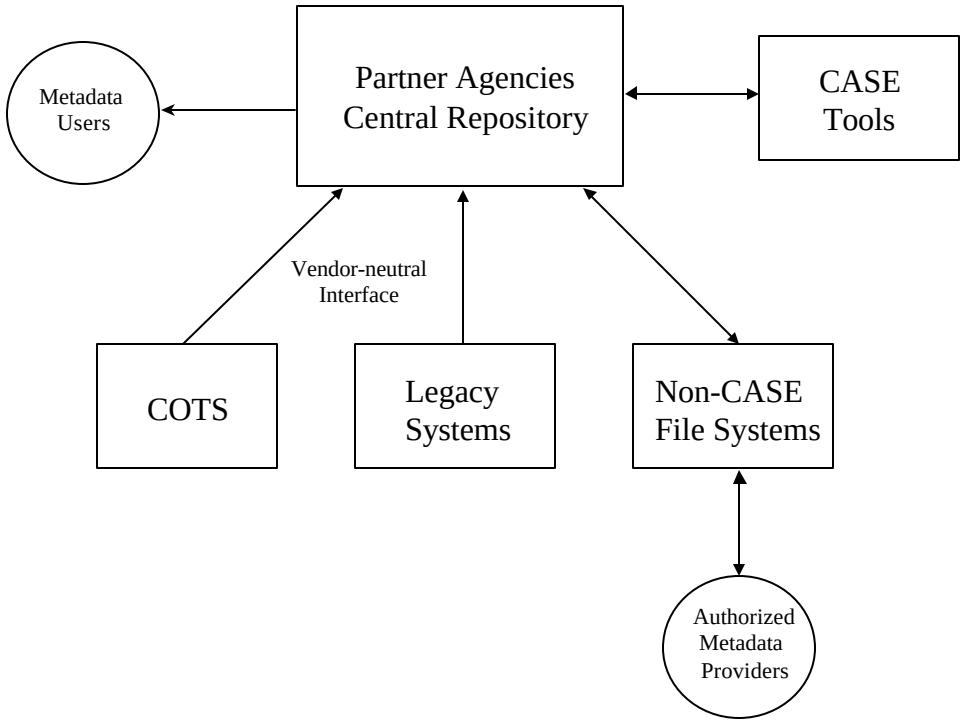

Exhibit 9 – Metadata Entered via Vendor-Neutral Interface

9.4 Metadata Standards (also listed in Section 4)

9.4.1 Metadata Content Standards

A metadata standard will be defined to facilitate consistent definition of the metadata. This standard will define a core set of metadata that is required, as well as optional metadata. The standard will be extensible to meet individual project needs. The standard will support relational, geospatial, multi-dimensional, document and other types of data. The standard will be consistent with industry standards.

9.4.2 Metadata Repository Interface Standards

An interface standard will be defined to facilitate transfer of information between the central repository and local CASE tools and other metadata sources. This standard will define the physical interface for metadata going in and coming out of the repository. The intent of this interface is to support an open architecture. The interface standard will identify those CASE tools for which "off-the-shelf" interfaces are available. The interface standard will also include a vendor-neutral interface format for metadata from other sources, such as COTS products, non-sanctioned CASE tools, and projects that don't use formal CASE tools.

9.4.3 Configuration Management/Quality Assurance Standards

Configuration Management standards will ensure the status of the information is properly and consistently identified and that prior versions of metadata can be maintained. Quality Assurance procedures will ensure entries are valid. Metadata providers are responsible for its accuracy. Security will be administered to ensure only authorized individuals perform updates. CM, QA and security procedures will control the addition and update of metadata.

9.5 Standard Data Elements

An ultimate objective for data sharing and data reuse is the establishing of "standard data elements." A standard data element is an element or structure which has a definition acknowledged by all partner agencies. The domain for standard elements may be slightly different for each use of the element. For example, customer name is a standard data element. It has a single definition, but the domain or actual set of customers may be different for each system or location where it is used.

Establishing standard elements usually requires a lengthy period of review by business areas in each of the partner agencies. When each agency has a slightly different definition for an element, negotiation of the definition may be required. In some cases, as for example in establishing a standard element called "common land unit," extensive meetings have been held with multi-agency representation. Because of the effort involved, candidates for standard elements are chosen where real advantage lies for the sponsoring business area. Once established, the ability to pass data between systems and agencies is greatly facilitated.

The metadata for standard elements is maintained in the central registry repository.

9.6 Managing Geospatial and Other Clearinghouse Metadata

The collection of clearinghouse metadata should occur at the same time new data items are created. For example, each time a new map is digitized, the related metadata should to be collected and stored with the map. GIS data is not self-describing, i.e. you cannot usually tell what a point, or line on a map represents without some type of label or legend or other descriptive information. Nor can you tell what area the whole map represents without identifying information such as latitude/longitude coordinates, known reference points, or political boundaries. Metadata is absolutely essential to maintaining spatial data, and must be collected and maintained diligently.

In many cases, geospatial metadata will need to be entered manually. The tools used to create geospatial data typically do not directly support collection of complete metadata. Some GIS vendors are beginning to provide a limited set of tools.

By executive order, geospatial metadata must be made available through network access to a clearinghouse repository. The cost of producing geospatial data is such that wide usage is imperative. Multiple clearinghouse repositories can be maintained by the agencies as long as they are pointed to by one central repository that becomes the primary entry point for querying metadata stores.

9.7 Other Repository Items

In addition to metadata, the partner agencies can also use the central repository to store standard reference lists. These include glossaries of business terms, standard abbreviation lists, and standard acronym lists.

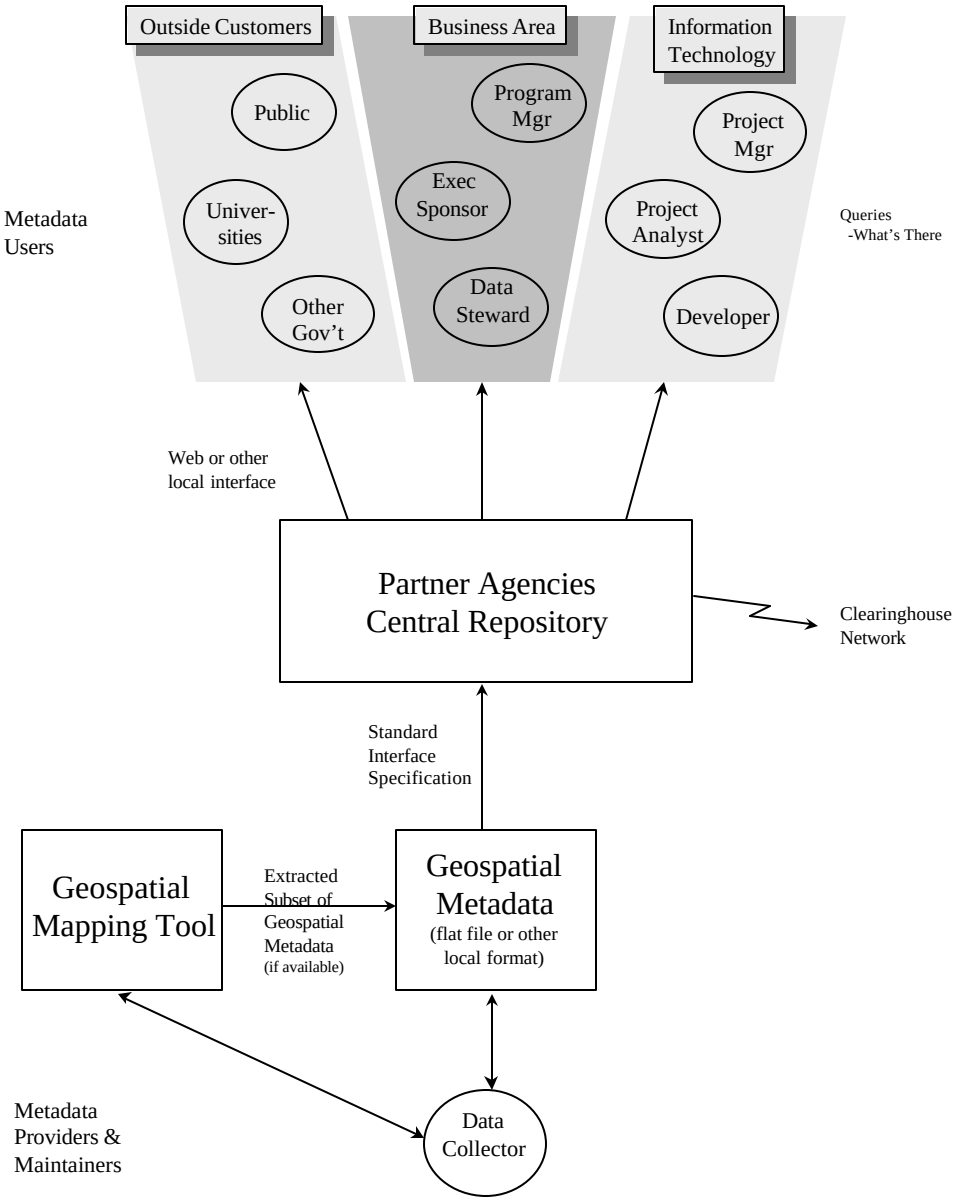

Exhibit 10 – Managing Geospatial Metadata

10 Appendix

10.1 Authorities and References

Authorities

a. Computer Security Act of 1987. (Public Law 100-235)
URL: http://csrc.nist.gov/secplcy/csa_87.txt

b. Departmental Regulation (DR) 3400-4 Departmental Data Administration Program.

c. Executive Order 12906, Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure, April 11, 1994.
URL: http://www.pub.whitehouse.gov/search/executive-orders.html

d. Office of Management and Budget (OMB) Circular A-130, Management of Federal Information Resources.
URL: http://www1.whitehouse.gov/WH/EOP/OMB/html/circular.html

e. OMB Circular A-16, Coordination of Surveying, Mapping and Related Spatial Data Activities, revised October 1990
URL: http://www1.whitehouse.gov/WH/EOP/OMB/html/circular.html

f. The Paperwork Reduction Act of 1980 (Public Law 96-511) as amended by the Paperwork Reduction Reauthorization Act of 1986.
URL: http://www.law.cornell.edu/uscode/44/3501.shtml

References

a. Charter of the Service Center Data Team.

b. Service Center Data Naming Standards, August 1998.

c. FGDC, Content Standards for Digital Geospatial Metadata, 1998.
URL: http://www.fgdc.gov/metadata/csdgm

d. Spatial Data Transfer Standard, FIPS 173

10.2 Acronyms

BIC	Business Integration Center
BPR	Business Process Reengineering
CASE	Computer Aided Software Engineering
CCC	Commodity Credit Corporation
CCE	Common Computing Environment
CD	Compact Disc
CDRS	Central Data Repository System
CIO	Chief Information Officers
CIP	Customer Information Profile
COTS	Commercial Off-the-Shelf
DATA TEAM	Service Center Data Team
DR	Departmental Regulation
DBMS	Database Management System
EIS	Executive Information System
FAC	Food and Agricultural Council
FGDC	Federal Geographic Data Committee
FSA	Farm Service Agency
FY	Fiscal Year
GILS	Government Information Locator System
GIS	Geographic Information Systems
GPRA	Government Performance and Results Act
I/O	Input/Output
IT	Information Technology
ITA	Information Technology Architecture
ITSB	Information Technology Services Branch
LAN/WAN	Local Area Network/Wide Area Network
L/W/V	LAN/WAN/Voice
MRB	Management Review Board
NRCS	Natural Resources Conservation Service
NSDI	National Spatial Data Infrastructure
OIP	Office Information Profile
OMB	Office of Management and Budget
PC	Personal Computer
RBS	Rural Business-Cooperative Service
RD	Rural Development

RUH	Rural Housing Service
RUS	Rural Utility Service
SCIT	Service Center Implementation Team
SSB	Support Service Bureau
TWG	Technical Working Groups
USDA	United States Department of Agriculture

10.3 Glossary

A more extensive and current glossary is maintained on the Service Center Data Team home page. Definitions on the web page may supercede the glossary contained in this document.

Business Rule

A statement that defines or constrains some aspect of the business as it is implemented in the data model. Data-related business rules are statements, phrased in absolute terms, about data (i.e., a telephone number must have 10 digits), and about relationships between data (i.e., if a phone number is entered, the phone type must also be entered.)

CASE Tools

Software tools used to automate software development activities.

Clearinghouse Metadata

Metadata that describes a physical instance of data that describes another set of data. Clearinghouse metadata is used to describe and locate physical instances of data or software. Clearinghouse metadata can describe relational databases, data files, documents, photographs, video segments, stored sounds, software systems, applications, software components, libraries, etc. The level of granularity of the metadata may vary. In one case, the metadata may describe actual data objects. In other cases, the metadata may contain pointers to other sources or repositories of metadata. One set of metadata may list all available photographs and where to find them. The metadata describing each photograph in detail may be stored elsewhere with the actual photograph.

Common Data

Data jointly owned, used, and managed by Service Center partners.

Component

A reusable software module that encapsulates a set of behavior, hiding implementation details.

Component Based Development

Creating applications by assembling existing components.

Configuration Item

A collection of hardware, software, and/or firmware, which satisfies an end-use function and is designated for configuration management. (IEEE 610.12-1990) Note: This is a very broad definition. Each category encompasses all the constituent items, such as metadata, data models, etc.

Configuration Management

A discipline applying technical and administrative direction and surveillance to: identify and document the functional and physical characteristics of a configuration item; audit the configuration items to verify conformance to specifications, interface control documents, and other contract requirements; control changes to configuration items and their related documentation; and record and report information needed to manage configuration items

effectively, including the status of proposed changes and the implementation of status of approved changes. (IEEE 610.12-1990)

Data

A discrete fact or value. Data is the raw material, which through its use and interpretation can provide valuable information. Data is the content of databases or data files.

Data Administration

The technical function of acquiring, defining, certifying, organizing, protecting, and delivering data and the metadata that describes it.

Data Administrator

The person who defines, organizes, manages, controls, protects, and standardizes data models, data elements, and other metadata.

Data Architecture

An orderly arrangement of Service Center data resources to achieve (1) a common understanding of data resources available; (2) a planned approach to data acquisition, storage, and retrieval to achieve a high degree of responsiveness to user demands; and (3) a high degree of data sharing and data mobility to reduce program delivery costs.

Data Clearinghouse

A facility for advertising and distributing datasets. Metadata describing available datasets is made available to a chosen audience (i.e., the general public, partner agencies, etc.). Packaged datasets (i.e., diskettes, CDs, etc.) or online retrieval of selected subsets of data are distributed within security guidelines. There may be a charge for obtaining datasets.

Data Dictionary

A database about data and database structures. A catalog of all data elements containing names structures and information about their usage. A central location for metadata. Normally, data dictionaries are designed to store a limited set of available metadata, concentrating on the information relating to the data elements, databases, files, and programs of implemented systems.

Data Integrity

The state that exists when data is handled as intended and is not exposed to accidental or malicious modification, destruction, or disclosure. Also, the preservation of data for its intended use.

Data Management

The managerial function of taking responsibility for data and the processes that support it. It focuses the strategic planning and operational data functions (technical planning, data administration, database administration, data warehouse administration) on meeting program delivery goals.

Data Mart

A type of data warehouse that contains smaller subsets of data and focuses on a particular business discipline or organizational component.

Data Model

A pictorial view of data, groupings of data, relationships between data groupings, or the organization of data groupings by dependencies. A "logical" data model is a view that does not depend on the characteristics of the computerized system or of the physical storage. A "physical" data model typically refines the logical model by adding the constraint incumbent to the database system or physical storage method.

Data Steward

A business area expert who is assigned responsibility for the data content of the database. The data steward establishes business rules, defines data elements, identifies valid data values, establishes certification standards, and establishes the completeness and availability of the data.

Data Storage Facility

Any national, state, or local location where data is stored and maintained.

Data Validation

Applying a set of rules, comparisons, or decisions to a data element to determine if it falls within the pre-established boundaries of values for that element.

Data Warehouse

An informational database, or collection of databases, used to store shareable data. The warehouse is usually created through data extracts from operational databases. The warehouse adheres to a single enterprise data model to ensure consistency of decision-support data across the enterprise. The warehouse typically allows users to tap into an organization's vast store of operational data to track and respond to business trends, and to facilitate forecasting and planning efforts.

Database

A collection of related data organized to serve one or more applications. In the broader sense, it describes any organized collection of data regardless of the physical storage method.

Database Administration

Encompasses the day-to-day technical functions that support ongoing business operations. It includes the collecting, defining, certifying, organizing, protecting, and delivery of both data and metadata (data about data).

Database Administrator

The person who creates, manages, controls, and protects a database.

Domain

A listing of all the valid values that can be stored in a data element.

Enterprise Data

All data owned and managed by all Service Center partners including common, shared and unique data.

Enterprise Data Architecture

See "Data Architecture." Emphasizes that the data architecture extends to the entire business enterprise of the partner Service Center agencies.

Enterprise Data Model

An overall pictorial view of the many applications and databases making up the participating agencies' combined data assets. The intent is to manage the overall data assets to achieve optimal integration, sharing, access, and utilization of technology resources and infrastructure.

Geospatial Data

Information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. This information may be derived from sources such as remote sensing, mapping, and surveying technologies. It includes both attributes (text) as well as spatial (map) data.

Geospatial (GIS) Metadata

Metadata that describes geospatial features. Points, lines, and polygons are created in a GIS tool and have little meaning if they are not described with words. These word descriptions can include such items as a name for the feature, category (i.e. farm, field, water, wetland), creation date, crop cover, ownership, etc. There is geospatial metadata, as well, that describes large sets of geospatial data. This includes metadata that describes a complete map, or a geospatial data layer that covers a large physical area. This type of geospatial metadata is used in particular to advertise or locate maps and geospatial datasets.

Information

A commodity derived from data through analysis or by the orderly presentation of data for human interpretation.

Local Database

A database that is used by an agency organization (e.g., region, state, center, institute, etc.) but does not meet the definition of a National Database. Examples include local spreadsheets, report extracts, and PC-based databases.

Metadata

Data about data. Metadata describes how, when, and by whom a particular set of data was collected, and how the data is formatted. Metadata includes attributes such as data name, length, domain of valid values, and definition. Metadata can also identify and describe a set of data or a complex data type such as a map, photograph, spatial data set, etc.

Metadata Repository

A database of information describing the characteristics (metadata) of data. Typically, the repository also stores a broad range of descriptive information, including business rules, data models, and process models that help to elaborate on the usage of data in various systems. Repositories can also store

metadata for the purpose of identifying and retrieving sets of actual data. Metadata that describes a map is an example.

National Database

A permanent database that (1) has international, national, USDA, or agency-wide application, (2) is included in a standard software suite, (3) contains data that is used/shared directly in making national program decisions, or (4) is used/shared in multiple offices, states, or other internal/external organizations.

Registry Metadata

Metadata that gives data elements a name and definition, describes how the data will be stored, records relationships between data items, captures the business rules for data element and entities, describes the domain of valid data values, and identifies the caretakers of the data. It may include data models and database designs for specific application(s) where the data is used.

Reverse Engineering

The development of a data model and associated specifications by analyzing existing databases and software systems by manual means, or through specialized software. This is usually done for applications that do not have a data model or specifications in existence or for which the model and specifications have become obsolete. The products of reverse engineering are used for maintenance, movement of the system to another platform, or provision of a baseline for re-design or re-engineering of the system.

Shared Data

Data owned and managed by a specific Service Center partner and shared by other partners.

Standard Data Element

An element or structure that has a definition acknowledged by all partner agencies.

Strategic Data Management Planning

Planning how data is acquired, stored, and used in the most efficient manner to support the agency mission.

Unique Data

Data owned and managed by a specific Service Center partner but not shared.

Warehouse Metadata

Metadata that describes the contents of a warehouse. This can include the original source of the data, timestamps, data conversion routines, data transformations, volatility of the data, refresh periods, data reliability indicators, relationships between data from multiple sources, etc. Warehouse metadata can also include metadata describing information and data elements stored in a warehouse, similar to registry and clearinghouse metadata.